Hi, my name is YuShuen. I’ll show you the detail of our method.

To determine the camera motion, we apply the SIFT algorithm to extract the corresponding features between neighboring frames. Based on the correspondence, we know where the current features move to in the next frame and thus, we are able to determine the camera motion. Specifically, there must be a rigid transformation that can approximately transform the features from the current frame to the features in the next frame. We consider this transformation as the camera motion.

By aligning the video frames, we can represent the video without the camera motion. In this example, the red points denote the background SIFT features and their positions are retained. Since the corresponding backgrounds are transformed to the same position, we are able to preserve the temporal coherence when resizing.

We represent each video frame using a coarse grid mesh and deform the meshes to achieve video resizing. To preserve the temporal coherence, our goal is to retain the corresponding pixels attaching together after the resizing. Due to the applied meshes are coarse, for each grid vertex, we determine the corresponding quad in the previous frame for it, and then preserve the relations of each vertex and its corresponding quad to roughly preserve the attachment of corresponding pixels.

Specifically, the projected vertex can be represented by the linear combination of its four corresponding vertices. We introduce an energy function to minimize the distance between the vertex and the linear interpolated position of the quad vertices after the resizing.

There must be some vertices that have no corresponding quads because the video frames are aligned based on the camera motion. For those vertices, we simply constrain the deformation of their Laplacian coordinates to be similar. Here, we didn’t constrain the temporally adjacent vertices to attach together because we solve for the deformed grid meshes under the aligned domain.

The above mentioned algorithm can achieve the temporal coherence of objects that are transformed to the same position. However, as we can see in this video, the man is still moving even though the video frames are aligned. This is because the object motions are independent to the camera motion. Thus, we need another constrains to preserve the foreground objects.

To preserve the consistent resizing of foreground objects, we detect the foreground moving area by computing the pixel color variations. Specifically, we generate a mosaic image by blending the aligned video frames and then compute the difference of pixel colors between the mosaic image and the aligned video frames. The regions with high color variations are considered as moving area and thus, we preserve the quads covering them to have the same resizing. In this example, the quads in different colors means that they belong to different moving areas and thus can be preserved respectively.

Here we show the results with and without our object motion preservation. As we can see, when the woman moves to different positions, she is resized differently without the object motion constraint. On the other hand, our object motion preservation successfully prevents this artifact.

In addition to the temporal constraints, we need the spatial content preservation to retain the prominent objects from distortion. Our temporal constraints can be combined with all the previous image warping methods since the temporal constraints are formulated into energy terms.

To minimize the visual distortions, we deform the video frames using our previous image resizing method, such that the distortions can be floated by all the available homogeneous regions. That is, we preserve the quad’s aspect ratio while allowing its size to be changed in order to float the distortion in multiple directions.

When preserving the prominent objects, the importance map is determined by the combination of the gradient magnitude and the Itti’s saliency measure. Since the foreground objects usually attract more attentions, we blend the aligned saliency maps to emphasize the importance of the moving area. Such that, we can better preserve the foreground objects from squeezing. As we can see in the first row, the high saliency values of the moving area are because the boat moves when the video frames are aligned.

By putting all the energy terms together, we minimize the objective function to achieve the spatial temporal preservation of video resizing. Although there are many unknown vertex positions, the performance of our system is still acceptable since the objective function is not highly non-linear. The performance depends on the number of vertex positions and as well the video contents. For the examples used in this paper, our system takes 2 to 5 frames per seconds in average when minimizing the objective function.

We show some results to demonstrate the effectiveness of our algorithm. The frame alignments are shown in the bottom left and the retargeted videos are shown in the bottom right.

We also compare our method with the linear interpolation, the improved seam carving, the content aware video resizing and the naïve extension of our previous image resizing method. The simple extension is to constrain the temporally adjacent vertices to move smoothly when the video is played. As we can see, the temporal artifacts introduced by the seam carving and the warping methods are different. This is because the seam carving method directly removes the interior pixels and the warping methods squeeze the objects by the linear interpolation. Clearly, only our method can successfully prevent the flickering and the waving artifacts since we take the temporal coherence issue into consideration.

It is known that the homogeneous scaling achieves the best temporal coherence when resizing videos since everything is applied with the same transformation. On the other hand, respectively resize each video frame without considering the temporal coherence can preserve the best aspect ratios of prominent objects. To minimize the visual artifacts, we compromise the objects’ aspect ratios and the temporal coherence to achieve the aim.

The video clips with different scenes can be resized independently because there are no corresponding objects inside. Preserving the temporal coherence of different scenes is not necessary. Therefore, the video clips are generally short and the whole sequence can be solved simultaneously. However, if there is a long video with the same scene, we can divide the video into short clips, resize respectively, and then merge them together to obtain the result. In this scenario, some overlapping areas are required to smooth the temporal incoherence between the neighboring clips.

Although our algorithm can prevent the flickering and the waving artifacts that are commonly introduced by the previous methods, there are still some spaces for the improvement. First, of course, our method would degenerate into linear interpolation if there are too many prominent objects in the video. Another problem is that we determine the camera motion based on the SIFT features and thus, the obtained camera motion might be unreliable if the backgrounds are homogenous. Finally, we treat each video frame as a plane and the interior pixels are repositioned by using a linear transformation. It means our method cannot well align the corresponding pixels at the same position if the pixel depths vary. This is because of the perspective projection, the objects that are close to the camera would move faster when the camera moves. In this case, our method would treat a part of backgrounds as foregrounds and make the system more constrained. Here is an example, which the backgrounds are homogeneous and are with various depths, our frame alignment fails since the obtained camera transformations are wrong. In this example, our system treats the lady and the homogeneous background regions as a large moving area and then degenerates into the linear interpolation.

In conclusion, we are the first members who introduced the temporal coherence preservation for video resizing. In addition to the interior spatial contents, we consider the camera and the object motions to achieve the consistent resizing of corresponding objects. Therefore, our content and motion aware technique can not only prevent the prominent objects from distortion but also can eliminate the flickering and the waving artifacts that are commonly produced by the previous methods.

Finally, we would like to thank the grants supporting this project. And also, thank you for your attention.